Exhibit A

Begault, Durand R., 3-D Sound for Virtual Reality and Multimedia, pages 172 – 176 (including at least pages 213 – 216 and Figure 5.12 of original text as cited) (1994)

these prefiltered sounds using the Focal Point Type 2 driver algorithm. For instance, to move from left 90 degrees to left 45 degrees, one makes a transisition, similar to amplitude panning, between the left 90-degree and the 0-degree position.

A Virtual Reality Application: NASA's VIEW System

The implementation of the Convolvotron 3-D audio device within NASA's VIEW system gives an example of how distributed 3-D audio devices and tone generators are integrated within a dedicated virtual reality application. Wenzel, *et al.* (1990) created a distributed audio system as part of a virtual environment for providing symbolic acoustic signals that are matched to particular actions in terms of the semantic content of the action taken. The sounds are not only spatialized in 3-D dimensions for conveying situational awareness information but are also interactively manipulated in terms of pitch, loudness and timbre.

Figure 5.11 shows the hardware configuration for the auditory display. The reality engine at the time consisted of a Hewlett Packard HP900/835 computer. The RS-232 port was used to drive a Digital Equipment DECTALK™ speech synthesizer, a Convolvotron, and two Ensoniq ESQ-M™ tone generators via an RS-232-MIDI converter. In the configuration, the Convolvotron receives data over the RS-232 port regarding the head-tracker orientation and action of a VPL Data Glove™, and spatializes the output of one of the two synthesizers. A standard audio mixer is used to sum the outputs of the synthesizer, Convolvotron, and speech synthesizer.

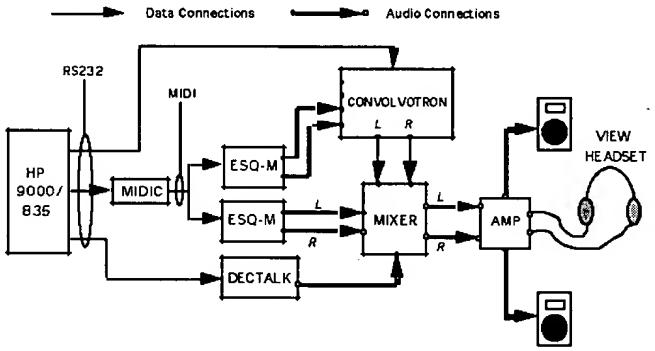

Figure 5.11. Implementation of the Convolvotron into the VIEW laboratory at NASA Ames Research Center (about 1987–1988). MIDIC is an RS-232 to MIDI converter. *Adapted from Wenzel,* et al., *1990.*

The Ensoniq ESQ-M tone generators were set to receive a variety of MIDI control information. Oscillator frequency, filter cut-off, and other parameters that are the building blocks of auditory icons can be mapped from the action of sensors that are input to the reality engine. While this type of configuration can be a formidable chore under normal circumstances, the VIEW audio subsystem was configured via specialized software (the auditory cue editor, ACE, developed by Phil Stone) that eased formation of the auditory symbologies and their connections to virtual world. In particular, the ability to do this off-line (in a stand-alone manner) from the reality engine is a practical consideration where availability of resources for development are frequently at a premium.

As already suggested earlier in this chapter, audio in any virtual reality configuration can provide an important source of feedback. For instance, the VPL Data Glove works by analyzing a combination of detected finger positions and then using the software to match them to sets of predefined gestures (e.g., a peace sign, pointing with the index finger, squeezing hand as a fist). It can be tricky at first to adapt one's hand actions to recognized gestures because individual variation in hand size and movement must match a generalized gesture. But if auditory feedback is supplied upon successful recognition, it aids the user in knowing when the action has been completed successfully. At Stanford Research Institute (SRI), Tom Piantanida and his colleagues supply auditory feedback when a fist gesture is recognized by playing a "squish" noise from a digital sampler that sounds similar to a wet washcloth being squeezed in the hand. This response is a more literal, or "representational," use of sound and might be termed an auditory icon since it is representative of an everyday sound (Gaver, 1986). By contrast, the VIEW project used a combination of two synthesized tones (e.g., the musical notes C and D) to give the user "glove state" feedback: An ascending sequence indicated a glove opening, and the reversed sequence indicated the glove closing. This type of "abstract" pitch sequence to which meaning is attached might be referred to as an nonrepresentational **earcon** (Blattner, Sumikawa, and Greenberg, 1989).

A common use of representative aural cues in virtual reality is as a means of replacing tactile sensation (haptic feedback) that would normally occur in a real environment. In virtual reality, one can walk through walls and objects; but by activating auditory information in the form of a bumping or crashing sound, the user receives feedback that a particular physical boundary has been violated. But after a while, using a simple crash for all encounters with different types of boundaries becomes increasingly unsatisfying, because it lacks complexity in response to human interaction. While the VIEW system had general application to many sorts of virtual environment scenarios and certainly could make use of these "representational" auditory cues, its use of sound for interaction in a **telepresence** context was vastly more sophisticated.

An example is the auditory "force-reflection" display used to substitute for a *range* of haptic (tactile) sensation involved in the teleoperated placement of a circuit card. Wearing a helmet-mounted display and data glove, the user was faced with the task of guiding a robot arm by watching the action within the virtual environment. Once the user's arm and hand were coupled with the robot equivalent and had grabbed the circuit card, the task was to correctly guide the card into a slot, a rather difficult undertaking with only visual feedback, given the latencies and complexity of visual displays of the time. In this scenario, **continuous variation** of sound parameters was used. If the card was being placed incorrectly, a basic soft, simple tone got louder, brighter, and more complex (via frequency modulation) the harder the user pushed to discourage damage to the misaligned card. This sound was accomplished by sending appropriate MIDI information on note volume and frequency modulation to the tone generator.

Another form of auditory feedback described by Wenzel, *et al.* (1990) was the use of **beat frequencies** between two tones to continuously inform the listener about the distance between the card and the slot. You may have noticed that a guitar being tuned produces a beating effect when two strings are close but not exactly in tune. As two identically tuned strings are slowly brought out of tune, the frequency of the beating increases. These beats are actually the perceived amplitude modulation of the summed waves, caused by the auditory system's inability to separate two frequencies on the basilar membrane smaller than the limit of discrimination (see Roederer, 1979). Using the beating as an acoustic range finder, the user could manipulate the card until it was successfully in the slot; the closer the card was to the slot, the closer the two tones came to the same frequency, with a corresponding reduction in the beat frequency. When the card was installed, the

tone stopped, the voice synthesizer would say "task completed," and a success melody was played (a quickly ascending scale, similar to the sound of a Macintosh starting up).

Computer Workstation 3-D Audio

A Generic Example

A simple audio equivalent to the graphic user interface (GUI) as described previously is only marginally useful. The concept of a computer workstation as not only a location for editing documents but also for **communicating** with databases and persons within the office adds a new dimension to the possibilities for 3-D sound. For instance, telephone communication can be handled though a computer via functions such as message recording and dialing. If the audio signal is brought into the computer and then spatialized, the communication source can be usefully arranged in virtual space. This feature becomes particularly important in multiconversation **teleconferencing**, where video images of the participants can be broadcast on the screen. Spatial audio teleconferencing has been described as an **audio window** system by Ludwig, Pincever, and Cohen (1990).

Consider all of the possible audio inputs to a workstation user, not only teleconferencing environments, but all types of sonic input. All types of sonic input could be directionalized to a specific location, controlled by the user. Furthermore, it is not necessary to place these audio sources in their corresponding visual locations; the audio spatial mapping can correspond to a **prioritization** scheme. For telephone calls, spatialization of incoming rings will become interesting when the caller can be identified. In such a scenario, calls from subordinates could be signaled by a phone ringing from the rear; a call from home could always ring from front right; and the intercom from the boss could be from front and center. In this case, spatial location informs the listener as to the prioritization for answering calls—whether or not to put someone on hold, or to activate a phone mail system.

Other types of input that could be spatialized to a headphone-wearing workstation user include building fire alarms, signals for sonification of the state of a particular computer process of machinery in another room, an intercom from a child's crib, as well as the more familiar "system beeps" that provide feedback from software such as word processors. And there's no reason why one's favorite CD cannot be placed in the background as well.

Figure 5.12 shows a hypothetical GUI for arranging these various sound inputs in virtual auditory space in a manner configurable by each individual listener. The basic concept is to represent each form of auditory input iconically on the screen and then allow manipulation of the distance and relative azimuth of each source. The computer interprets the relative location of the icons on the screen as commands to a multiinput, 3-D sound spatialization subsystem.

In noisy environments, the 3-D audio workstation system could also include **active noise cancellation**. A microphone placed on the outside of the headphones is used with a simple phase-inversion circuit to mix in a 180-degree inverted version of the noise with the desired signal at the headset. In reality, active noise cancellation headphones cannot completely eliminate all noises; most models provide about a 10 dB reduction in steady-state sound frequencies below 1 kHz. But when no signal is applied to the headphones except the inverse noise signal, it is surprising how well active noise cancellation headphones work in a typical office environment. In informal experiments by the author using lightweight Sennheiser HDC-450 NoiseGard® mobile headsets, the sound of HVAC systems, whirring computer fans, and outdoor traffic is noticeably reduced. The potential for using active noise cancellation in workstation 3-D audio systems seems great.

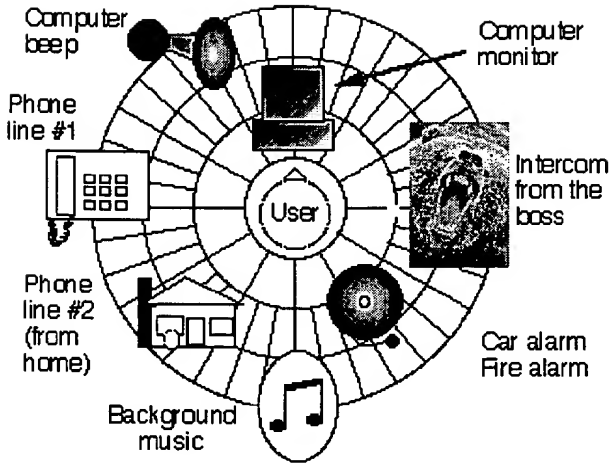

Figure 5.12. Layout for a hypothetical GUI for arranging a set of incoming sounds. By resizing the icons, volume could be adjusted independently of perceived distance. *Graphic assistance by Denise Begault.*

Audio Windows

The concept of **audio windows** has been explored extensively at the Human Interface Laboratory of Nippon Telegraph and Telephone Corporation by Cohen and his associates (Cohen and Koizumi, 1991; Cohen, 1993). They have developed an experimental glove-based system called "handysound" and a more viable interface called MAW (Multidimensional Audio Windows) that uses 3-D sound to control the apparent azimuthal location of conference participants. A user of the MAW system sees a display of circular icons with pictures of each participant from a bird's eye perspective, arranged about a virtual desk. The visual arrangement of icons translates into instructions for a 3-D audio subsystem.

The gain and azimuth heard for each icon is linked to the relative position on the graphic display; a mouse can be used to make sources louder by resizing the icon. MAW allows teleconference participants to wander around in the virtual room and to focus their voice toward particular conference participants by using a binaural HRTF to process each virtual source. Distance cues also are modeled in the system, and one can leapfrog between teleconference sessions or be at several simultaneously simply by cutting and pasting one's personal graphic icon to different environments, each represented by a separate window. One can also move to a "private conference" by moving the icons of the participants to a special region on the computer screen, symbolized as a separate room.

Audio GUIs for the Blind

Burgess (1992) and Crispien and Petrie (1993) have described interesting applications of spatial sound for allowing blind persons to navigate within a GUI. Crispien and Petrie's concept is applicable to either a headphone-based 3-D sound system or a multiple loudspeaker system. Auditory icons were interfaced with a Microsoft® Windows™ operating system using the icons to respond to various GUI interactions (see Table 5.2). These sounds were then spatialized according to the position on the computer screen by a two-dimensional auditory display (azimuth and elevation). Because GUIs are not friendly operating systems for blind persons, development of spatialized

acoustic cues might have great potential. But presently there are no studies comparing blind people's non-GUI system performance versus a 3-D audio cued, GUI system where subjects are given equivalent periods of training and practice on both systems.

Table 5.2. Some of the spatialized auditory icons used by Crispien and Petrie (1993) in their GUIB (Graphical User Interface for the Blind) system. Interactions are actions taken by the user on the GUI; auditory icons, a term coined by Gaver (1986) are sounded in response to actions on the screen and then spatialized to the appropriate position. Compare to Gaver's (1989) *The Sonic Finder: An Interface That Uses Auditory Icons.*

Interactions	Auditory icons
mouse tracking	steps walking
window pop-up	door open
window moving	scratching
window sizing	elastic band
buttons	switches
menu pop-up	window shade

Recording, Broadcasting , and Entertainment Applications

Loudspeakers and Cross-talk Cancellation

Many people want to control 3-D sound imagery over **loudspeakers** as opposed to headphones (see the section in Chapter 1, "Surround versus 3-D Sound"). Most have experienced "supernormal" spatial experiences with normal stereo recordings played over stereo loudspeakers, and there are countless methods to produce exciting spatial effects over loudspeakers that require nothing of current 3-D sound technology. And there have been and probably always will be add-on devices for creating spatial effects either in the recording studio or in the home (one of the oldest tricks involves wiring a third in-between speaker out-of-phase from the stereo amplifier leads). The problem is that it's difficult to harness 3-D spatial imagery over loudspeakers in such a way that the imagery can be transported to a number of listeners in a predictable or even meaningful manner.

There are three reasons why 3-D sound over loudspeakers is problematic, compared to headphone listening.

1. The environmental context of the listening space, in the form of reflected energy, will always be superimposed upon the incoming signal to the eardrums, often with degradative or unpredictable effects. This distortion is especially true for strong early reflections within 10–15 msec, although these can be minimized by listening in a relatively acoustically damped room or in an anechoic chamber.
2. It is impossible to predict the position of the listener or of the speakers in any given situation and impossible to compensate for multiple listeners. One can go to extremes to put people in what is known as the **sweet spot**, best defined as the location closest to where the original